Data Visualizations: 
Drawing Actionable Insights From Data

Katy Börner

Victor H. Yngve Professor of Information Science 
Director, Cyberinfrastructure for Network Science Center 
School of Informatics and Computing and 
Indiana University Network Science Institute 
Indiana University, USA

Inaugural Lecture 
Auditorium A, CDC Tom Harkin Global Comm. Center, Atlanta, GA

February 4, 2016

scimaps.org
How can we communicate the beauty, structure, and dynamics of science to a general audience?
Debut of 5th Iteration of the Mapping Science Exhibit at MEDIA X was in 2009 at Wallenberg Hall, Stanford University.

Science Maps in “Expediton Zukunft” science train visited 62 cities in 7 months. Opening on April 23rd, 2009 by German Chancellor Merkel
Ingo Gunther's Worldprocessor globe design on display at the Museum of Emerging Science and Innovation in Tokyo, Japan.

Places & Spaces Digital Display in North Carolina State’s Immersion Theater
Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA
January 25-June 17, 2016
10 iterations over 10 years
equal

10 \times 10 = 100 \text{ maps!}
Cartographic maps of physical places have guided mankind’s explorations for centuries.

They enabled the discovery of new worlds while also marking territories inhabited by the unknown.

Without maps, we would be lost.

Cosmographia World Map - Claudius Ptolemy - 1482
Science maps of abstract semantic spaces aim to serve today’s explorers navigating the world of science.

They can be used to identify objectively major experts, institutions, collections. They allow us to track the emergence, evolution, and disappearance of topics and help to identify the most promising areas of research.
In Terms of Geography - Andre Skupin - 2005

The Structure of Science - Kevin Boyack, Richard Klavans - 2005
How would a reference system for all of science look?

What dimensions would it have?
HistCite™ Visualization of DNA Development - Eugene Garfield, Elisha Hardy, Katy Börner, Ludmila Pollock, Jan Witkowski - 2006

Taxonomy Visualization of Patent Data - Katy Borner, Elisha Hardy, Bruce Herr, Todd Holloway, Bradford Paley - 2006
TexArc Visualization of “The History of Science” - W. Bradford Paley - 2006

The Power of Forecasts 2007
Impact of Air Travel on Global Spread of Infectious Diseases - Vittoria Colizza, Alessandro Vespignani - 2007
Can one forecast science?

What ‘science forecast language’ will work to communicate results?
114 Years of Physical Review - Bruce W. Herr II, Russell Duhon, Katy Borner, Elisha Hardy, Shashikant Penumarthi - 2007

What insight needs to economic decision makers have?

What data views are most useful?
Europe Raw Cotton Imports in 1858, 1864 and 1865 - Charles Joseph Minard - 1866

Tracing of Key Events in the Development of the Video Tape Recorder - Mr. G. Benn, Francis Narin - 1968
Chemical Research & Development
Powers the U.S. Innovation Engine

INVESTMENT IN CHEMICAL SCIENCE R&D

FEDERAL GOVERNMENT

$1 Billion
FEDERAL FUNDING

$5 Billion
INDUSTRY FUNDING

$8 Billion
TAXES

CHEMICAL INDUSTRY

$1B
3 YEARS
INFORMATION RESEARCH

$1B + $5B
10 YEARS
INDUSTRY COMMERCIALIZATION

$10 Billion
CHEMICAL INDUSTRY OPERATING INCOME

$40 Billion
GROWTH IN GNP
+ 600,000
JOBS CREATED

U.S. ECONOMY

The design shows that an input of
$1B in federal investment, leveraged
by $1B in industry investment, brings
new technologies to market and
results in $40B of economic income
for the chemical industry. $40B
growth in the Gross National Product
GNP and further impacts the US
economy by generating approximately
600,000 jobs, along with a return of
$40B in taxes. Additional results, also
reported in the CCRI studies, are
described in the map to the left. This
map clearly shows the two R&D
investment cycles: the shorter
industry’s investment at the innovation
stage to commercialization cycle; and
the longer federal investment cycle
which begins at basic research and
outlines in national economic and
job growth along with the increased tax
base that in turn is available for
investment in basic research.
Science Maps for Scholars 2010
The Emergence of Nanotechnology - Loet Leydesdorff - 2010

Science Maps as Visual Interfaces to Digital Libraries 2011
Map of Scientific Collaborations from 2005-2009

Stream of Scientific Collaborations between World Cities - Olivier H. Beauchesne - 2012

History of Science Fiction - Ward Shelley - 2011
Pulse of the Nation - Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Niels Rosenquist - 2010
Who Really Matters in the World—Leadership Networks in Different-Language Wikipedias
Peter A. Gloor, Keiichi Nemoto, Samuel T. Mills, and David E. Polley - 2013

The Future of Science Mapping 2014
Use the original online tool at healthmap.org/predict


*Map of the Internet* - Martin Vargic - 2014
Exploring the Relationships between a Map of Altruism and a Map of Science - Richard Klavans and Kevin W. Boyack - 2014

Explore the maps and background information at http://scimaps.org

**Background and Goals**

The Places & Spaces: Mapping Science exhibit was created to communicate human activity and scientific progress on a plot that enable the close inspection of large-scale maps in public conferences; (2) novel, interactive macroscope tools that let Themes for the upcoming iterations/years are:

- 11th Iteration (2013): Macroscopes for Interacting With Science
- 12th Iteration (2014): Macroscopes for Making Sense of Science
- 14th Iteration (2016): Macroscopes for Economic Decision Makers

**Microscopes, Telescopes, Macroscopes Plug-and-Play Macroscopes**

**The Infinitely Great**

![Telescope](image1.png)

**The Infinitely Small**

![Microscope](image2.png)
MACROSCOPES FOR INTERACTING WITH SCIENCE

http://scimaps.org/iteration/11

Visit us on Facebook!

Become a fan and see many great photos of the exhibit—plus find out when it's coming to a venue near you!

facebook.com/mappingscience
Data Visualizations:
Drawing Actionable Insights From Data

Katy Börner

Victor H. Yngve Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing and
Indiana University Network Science Institute
Indiana University, USA

Inaugural Lecture
Auditorium A, CDC Tom Harkin Global Comm. Center, Atlanta, GA

February 4, 2016
How to Classify Different Visualizations?

By
• User insight needs?
• User task types?

• Data to be visualized?
• Data transformation?

• Visualization technique?
• Visual mapping transformation?
• Interaction techniques?

• Or?
Different Question Types

Find your way

Descriptive & Predictive Models

Find collaborators, friends

Identify trends

Terabytes of data

Different Levels of Abstraction/Analysis

Macro/Global
Population Level

Meso/Local
Group Level

Micro
Individual Level
### Tasks

<table>
<thead>
<tr>
<th>LEVELS</th>
<th>MICRO: Individual Level</th>
<th>MESO: Local Level</th>
<th>MACRO: Global Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>about 3–1,000 records</td>
<td>about 1,001–100,000 records</td>
<td>more than 100,000 records</td>
</tr>
</tbody>
</table>

### Needs-Driven Workflow Design

**Stakeholders**

- **Types and levels of analysis** determine data, algorithms, and parameters, and deployment

**Data**

- **READ**
- **ANALYZE**

**DEPLOY**

- **Visually encode data**
- **Overlay data**
- **Select visualiz. type**
Needs-Driven Workflow Design

Stakeholders

Types and levels of analysis determine data, algorithms & parameters, and deployment

Data

READ

ANALYZE

DEPLOY

Validation

Interpretation

Visually encode data

Overlay data

Select visualiz. type

VIEW

Graphic Variable Types

Modify reference system, add records & links

Visualization Types (reference systems)

82
Visualization Types (Reference Systems)

1. **Charts:** No reference system—e.g., Wordle.com, pie charts

2. **Tables:** Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.

3. **Graphs:** Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.

4. **Geospatial maps:** Use latitude and longitude reference system. World or city maps.

5. **Network graphs:** Node position might depends on node attributes or node similarity. **Tree graphs:** hierarchies, taxonomies, genealogies. **Networks:** social networks, migration flows.

IVMOOC App – More than 60 visualizations

The “IVMOOC Flashcards” app can be downloaded from Google Play and Apple iOS stores.
Visualization Framework

<table>
<thead>
<tr>
<th>Insight Need Types</th>
<th>Data Scale Types</th>
<th>Visualization Types</th>
<th>Graphic Symbol Types</th>
<th>Graphic Variable Types</th>
<th>Interaction Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>page 26</td>
<td>page 28</td>
<td>page 30</td>
<td>page 32</td>
<td>page 34</td>
<td>page 26</td>
</tr>
<tr>
<td>categories/clusters</td>
<td>nominal</td>
<td>table</td>
<td>geometric symbols</td>
<td>spatial position</td>
<td></td>
</tr>
<tr>
<td>order/rank/sort</td>
<td>ordinal</td>
<td>chart</td>
<td>point</td>
<td>retinal form</td>
<td></td>
</tr>
<tr>
<td>distributions (also outliers, gaps)</td>
<td>interval</td>
<td>graph</td>
<td>line</td>
<td>color</td>
<td></td>
</tr>
<tr>
<td>comparisons</td>
<td>ratio</td>
<td>map</td>
<td>area</td>
<td>optics</td>
<td></td>
</tr>
<tr>
<td>trends (process and time)</td>
<td>geospatial</td>
<td>network layout</td>
<td>surface</td>
<td>motion</td>
<td></td>
</tr>
<tr>
<td>compositions (also of text)</td>
<td>correlations/relationships</td>
<td></td>
<td>volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linguistic symbols</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>text</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>numerals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>punctuation marks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pictorial symbols</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>images</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>icons</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>statistical glyphs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

See page 24

Visualization Framework

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td>categorize</td>
<td>category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>rank</td>
<td>ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution</td>
<td>distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compare</td>
<td>nominal</td>
<td>differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison &amp; deviation</td>
<td>time series</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>patterns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>over time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>process and time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>geospatial</td>
<td>spatial</td>
<td>location</td>
<td></td>
<td></td>
<td>generate maps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>part-to-whole</td>
<td>proportions</td>
<td></td>
<td></td>
<td>see parts of whole, analyze text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>form and structure</td>
<td></td>
<td></td>
<td>composition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(also of text)</td>
<td></td>
<td></td>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>association</td>
<td>correlate</td>
<td>correlation</td>
<td></td>
<td></td>
<td>relations between data points</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>hierarchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

86
Visualization Framework

<table>
<thead>
<tr>
<th>Insight Need Types</th>
<th>Data Scale Types</th>
<th>Visualization Types</th>
<th>Graphic Symbol Types</th>
<th>Graphic Variable Types</th>
<th>Interaction Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>categorize/cluster</td>
<td>nominal</td>
<td>table</td>
<td>geometric symbols</td>
<td>spatial</td>
<td>overview</td>
</tr>
<tr>
<td>order/rank/sort</td>
<td>ordinal</td>
<td>chart</td>
<td>point</td>
<td>position</td>
<td>zoom</td>
</tr>
<tr>
<td>distributions</td>
<td>interval</td>
<td>graph</td>
<td>line</td>
<td>retinal form</td>
<td>search and locate</td>
</tr>
<tr>
<td>(also, outliers, gaps)</td>
<td>ratio</td>
<td>map</td>
<td>area surface</td>
<td>color</td>
<td>filter</td>
</tr>
<tr>
<td>comparisons</td>
<td></td>
<td>network layout</td>
<td>surface</td>
<td>optics</td>
<td>details-on-demand</td>
</tr>
<tr>
<td>trends</td>
<td></td>
<td></td>
<td>volume</td>
<td>motion</td>
<td>history</td>
</tr>
<tr>
<td>(process and time)</td>
<td></td>
<td></td>
<td>linguistic symbols</td>
<td>extract</td>
<td>extract</td>
</tr>
<tr>
<td>geospatial</td>
<td></td>
<td></td>
<td>text</td>
<td>link and brush</td>
<td>projection</td>
</tr>
<tr>
<td>compositions</td>
<td></td>
<td></td>
<td>numerals</td>
<td>projection</td>
<td>distortion</td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td>punctuation marks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>correlations/relationships</td>
<td></td>
<td></td>
<td>pictorial symbols</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>images</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>icons</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>statistical glyphs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graphic Variable Types Versus Graphic Symbol Types

<table>
<thead>
<tr>
<th>Spatial</th>
<th>Geometric Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>Point</td>
</tr>
<tr>
<td>y</td>
<td>Line</td>
</tr>
<tr>
<td>z</td>
<td>Area</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size</th>
<th>Quantitative</th>
<th>NA (Not Applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>Qualitative</td>
<td>NA</td>
</tr>
<tr>
<td>Rotation</td>
<td>Quantitative</td>
<td>NA</td>
</tr>
<tr>
<td>Curvature</td>
<td>Quantitative</td>
<td>NA</td>
</tr>
<tr>
<td>Angle</td>
<td>Quantitative</td>
<td>NA</td>
</tr>
<tr>
<td>Class</td>
<td>Qualitative</td>
<td>NA</td>
</tr>
<tr>
<td>Value</td>
<td>Quantitative</td>
<td></td>
</tr>
<tr>
<td>Hue</td>
<td>Qualitative</td>
<td></td>
</tr>
<tr>
<td>Saturation</td>
<td>Quantitative</td>
<td></td>
</tr>
</tbody>
</table>
Load **One** File and Run **Many** Analyses and Visualizations

<table>
<thead>
<tr>
<th>Times Cited</th>
<th>Publication Year</th>
<th>City of Publisher</th>
<th>Country</th>
<th>Journal Title (Full)</th>
<th>Title</th>
<th>Subject Category</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2011</td>
<td>NEW YORK</td>
<td>USA</td>
<td>COMMUNICATIONS OF THE ACM</td>
<td>Plug-and-Play Macrosopes</td>
<td>Computer Science</td>
<td>Borner, K</td>
</tr>
<tr>
<td>18</td>
<td>2010</td>
<td>MALDEN</td>
<td>USA</td>
<td>CTS - CLINICAL AND TRANSLATIONAL SCIENCE</td>
<td>Advancing the Science of Team Science</td>
<td>Research &amp; Experimental Medicine</td>
<td>Falk-Krzewinski, HJ; Borner, K; Contractor, N; Fiore, S; Hall, KL; Keyton, J; Spring, B; Stokols, D; Trochim, W; Uzzi, B</td>
</tr>
<tr>
<td>13</td>
<td>2010</td>
<td>WASHINGTON</td>
<td>USA</td>
<td>SCIENCE TRANSLATIONAL MEDICINE</td>
<td>A Multi-Level Systems Perspective for the Science of Team Science</td>
<td>Cell Biology</td>
<td>Research &amp; Experimental Medicine</td>
</tr>
</tbody>
</table>
Load **One** File and Run **Many** Analyses and Visualizations

<table>
<thead>
<tr>
<th>Times Cited</th>
<th>Publication Year</th>
<th>City of Publisher</th>
<th>Country</th>
<th>Journal Title (Full)</th>
<th>Title</th>
<th>Subject Category</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2011</td>
<td>NEW YORK</td>
<td>USA</td>
<td>COMMUNICATIONS OF THE ACM</td>
<td>Plug-and-Play Macrosopes</td>
<td>Computer Science</td>
<td>Borner, K</td>
</tr>
</tbody>
</table>

Co-author and many other bi-modal networks.

Course Schedule

Part 1: Theory and Hands-On
• Session 1 – Workflow Design and Visualization Framework
• Session 2 – “When:” Temporal Data
• Session 3 – “Where:” Geospatial Data
• Session 4 – “What:” Topical Data

Mid-Term
• Session 5 – “With Whom:” Trees
• Session 6 – “With Whom:” Networks
• Session 7 – Dynamic Visualizations and Deployment

Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).
CDC Tutorial on Friday Feb 5, 2016, 8:30-11:30am

Title: Open Source Tools for Data Analysis and Visualization  
Speaker: Prof. Katy Börner, Indiana University

Abstract: This tutorial is designed for researchers and practitioners interested to use advanced data mining algorithms and visualizations in their research and daily decision making. It introduces the open source Science of Science (Sci2) Tool that supports temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels. Open data from different government agencies will be used to demonstrate different analysis and visualization workflows.

The tutorial provides “hands-on” training. Please bring your laptop and pre-install the Sci2 (v 1.1 beta) tool prior to the workshop.

Additional theory and hands-on lectures are available in the Information Visualization MOOC (IVMOOC) (http://ivmooc.cns.iu.edu) that is taught each Spring for students from 100+ countries.
This conference is funded by the NSF Science of Science and Innovation Policy (SciSIP) program and aims at facilitating the generation and execution of a new Roadmap for the Science of Science Policy community and a strategic plan for SciSIP program, see details at http://modsti.cns.iu.edu.

Every 10 years the OECD Blue Sky Forum engages the policy community, data users and providers into an open dialogue to review and develop its long-term agenda on science, technology and innovation (STI) data and indicators, see details at http://www.oecd.org/science/blue-sky.htm
References


All papers, maps, tools, talks, press are linked from http://cns.iu.edu

These slides are at http://cns.iu.edu/docs/presentations

CNS Facebook: http://www.facebook.com/cnscenter

Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience